УДК 004.8:004.9

Я.М. Карандашев, Б.В. Крыжановский
Центр оптико-нейронных технологий НИИСИ РАН, г. Москва, Россия
iont.niisi@gmail.com

Эффективное увеличение области притяжения глобального минимума квадратичного бинарного функционала при нейросетевом поиске*

Решается задача минимизации квадратичного функционала в конфигурационном пространстве. Для эффективного увеличения области притяжения глубоких минимумов предлагается матрицу, на которой построен функционал, возводить в степень, и на полученном новом функционале решать задачу минимизации. В работе показано на примере матриц двумерной спинстекольной модели Изинга, что такая техника приводит к сдвигу спектра минимумов в более глубокую область, резко сокращает число находимых мелких минимумов и позволяет с большей, на 3 – 4 порядка, вероятностью находить глобальный минимум.

Введение

В данной работе мы рассматриваем задачу нахождения глобального минимума квадратичного функционала в конфигурационном пространстве.

Пусть T – вещественная матрица размера $N \times N$, симметричная и с нулевой диагональю. Квадратичный функционал, построенный на этой матрице, имеет вид:

$$E_1 = -\frac{1}{2}\sum_{i=1}^{N}\sum_{j}^{N} T_{ij} s_i s_j, \; s_{i,j} = \pm 1. \qquad (1)$$

Задача состоит в том, чтобы найти такой конфигурационный вектор $S_0 = (s_1^0, s_2^0, ..., s_N^0)$, который даёт глобальный минимум функционала.

Данная задача является NP-полной. Её можно решить, воспользовавшись нейросетевым спуском со случайных стартов. Пусть у нас есть некоторая начальная (вообще говоря, случайная) конфигурация спинов $S = (s_1, s_2, ..., s_N)$. Найдём локальное поле, действующее на каждый её спин:

$$h_i = -\partial E_1 / \partial s_i = \sum_{j \neq i} T_{ij} s_j. \qquad (2)$$

Последовательно разворачивая спины так, чтобы они совпадали по знаку с действующим на них локальным полем, энергия конфигурации будет понижаться до тех пор, пока мы не застрянем в одном из локальных минимумов, где все спины будут направлены вдоль действующего на них локального поля. Чтобы найти глобальный минимум, данный спуск придётся провести многократно с различных стартовых конфигураций. На самом деле нейросетевой спуск является аналогом покоординатного спуска в вещественном пространстве.

* Выполнена в рамках Гранта Президента Российской Федерации по государственной поддержке ведущих научных школ, номер гранта НШ – 365.2008.9

Исходные данные

В нашей работе мы будем иметь дело лишь с функционалами, построенными на матрицах спинового стекла двумерной модели Изинга. Модель Изинга – это модель взаимодействия спинов, находящихся в узлах решётки и взаимодействующих только с ближайшими соседями. Рассматриваемая размерность задачи $N = 100$.

Данный тип матриц выбран по нескольким причинам:

Во-первых, для них описанный выше обычный случайный нейросетевой поиск находит глобальный минимум менее чем один раз за миллион стартов. При этом находится порядка 990 тысяч других (более мелких) минимумов. Хотелось бы найти лучший способ поиска.

Во-вторых, для рассматриваемых матриц Изинга каждой конфигурации S соответствует ортогональная ей конфигурация S' ($SS' = 0$) с противоположной по знаку энергией $E_1(S') = -E_1(S)$. Этот факт окажется полезным в дальнейшем.

И, наконец, в-третьих, в работе [1] показано, как с помощью branch and cut метода можно найти глобальный минимум функционала на рассматриваемых матрицах небольших размерностей, используя их сильную разреженность. Значит, в руках у нас уже есть конфигурации глобальных минимумов, что помогает нам при оценке результативности наших методов.

Предлагаемый алгоритм

Как показано в работе [2], в обобщённой модели Хопфилда при нейросетевом поиске вероятность попадания в минимум тем больше, чем больше глубина минимума. В связи с этим наша базовая идея состоит в том, чтобы видоизменить энергетическую поверхность функционала (1) таким образом, чтобы его глубокие минимумы (которые мы ищем) стали ещё глубже, а значит, находились бы с большей вероятностью, при этом мелкие минимумы (которые для нас не представляют интереса) стали мельче или совсем исчезли из виду.

Покажем, как возведение исходной матрицы в квадрат реализует эту идею. Для этого представим нашу симметричную матрицу в виде взвешенного квазихеббовского разложения по внешним произведениям конфигурационных векторов:

$$T = \sum_{m=1}^{M} r_m S_m S_m^+, \tag{3}$$

где S_m – некоторые конфигурации, r_m – веса этих конфигураций, а M – число, достаточное для разложения. В работе [3] показано, что в качестве конфигураций в разложении (3) можно взять конфигурации экстремумов (в том числе глобальных максимумов и минимумов). В этом случае соответствующие им веса с точностью до некоторых флуктуаций будут пропорциональны энергиям этих экстремумов.

С учётом выражения (3) квадрат матрицы T примет вид:

$$T^2 = A + R, \tag{4}$$

где

$$A = N \sum_{m}^{M} r_m^2 S_m S_m^+ , \quad R = \sum_{m=1}^{M} \sum_{n=1}^{N} (1 - \delta_{mn}) r_m r_n S_m S_n (S_m, S_n) . \tag{5}$$

Первое из слагаемых (матрица A) с точностью до множителя N (который уйдёт при соответствующей нормировке) совпадает с исходным разложением матрицы (3), при этом веса соответствующих конфигураций оказались возведёнными в квадрат. Именно матрица A даёт нам функционал с нужным образом видоизменённой энергетической поверхностью. Действительно, большие веса при возведении в квадрат станут ещё больше и, следовательно, увеличится глубина соответствующих минимумов и вероятность их нахождения. Малые веса при возведении в квадрат станут еще меньше, и мелкие минимумы уйдут из рассмотрения.

Нельзя забывать и о втором (перекрёстном) члене R. Если бы конфигурации в разложении (3) были ортогональными, то мы бы имели $R = 0$, и новый функционал, построенный на квадрате матрицы:

$$E_2 = -\frac{1}{2}\sum_{i=1}^{N}\sum_{\substack{j=1\\ j\neq i}}^{N}(T^2)_{ij}\, s_i s_j, \qquad (6)$$

имел бы конфигурации локальных минимумов, совпадающие с конфигурациями исходного функционала. Однако в общем случае это неверно. Среднее значение элементов матрицы R равно нулю, а их стандарт порядка $1/\sqrt{N}$. Это означает, что вклад матрицы R в E_2 относительно невелик. Однако его наличие приводит к тому, что конфигурации локальных минимумов функционала E_2 слегка сдвигаются относительно конфигураций S_m исходного функционала E_1.

Полученные результаты

На рис. 1 приведены усреднённые по 50 матрицам графики спектральной плотности минимумов функционалов E_1 и E_2. Во всех экспериментах мы нормировали каждую матрицу (как исходную T, так и матрицу-квадрат $T^2 = T^+T$) на дисперсию её элементов. Как видно из рис. 1, спектр функционала E_2 сдвинут влево по сравнению со спектром исходного функционала E_1. Также на рис. 1 показано, куда сдвинулся глобальный минимум исходного функционала. Он стал заметно глубже. Число обнаруженных за миллион стартов минимумов функционала E_2 порядка 600 тысяч, что существенно меньше, чем для исходного (около 990 тысяч). Это показывает, что мы добились желаемого: глубокие минимумы стали глубже, а многие мелкие исчезли.

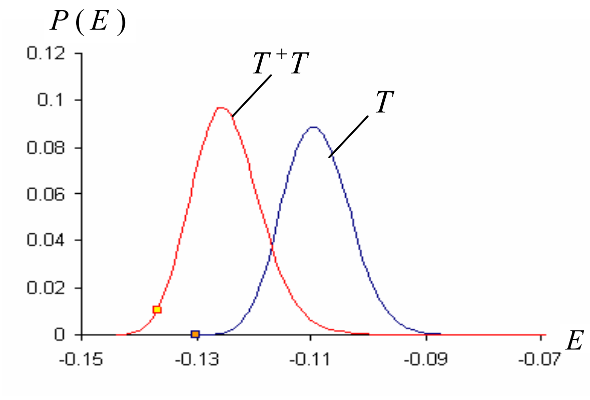


Рисунок 1 – График спектральной плотности минимумов функционала для матрицы Изинга (Т) и квадрата этой матрицы (Т*Т). Квадратами показано изменение энергии конфигурации глобального минимума матрицы Т

В случаях, если конфигурация глобального минимума исходного функционала не являлась локальным минимумом функционала E_2, мы брали энергию минимума, ближайшего к глобальному по расстоянию (число несовпадающих спинов). Число отличающихся спинов расположилось в диапазоне от 0 до 6, и в среднем по 50 матрицам оказалось равным 3,4, что подтверждает малость члена R в разложении (4).

Другим интересным подтверждением того, что при возведении матрицы в квадрат веса конфигураций тоже возводятся в квадрат, является следующее. Как упоминалось выше, спектр матрицы Изинга обладает тем свойством, что произвольному локальному минимуму функционала, построенного на этой матрице, соответствует некоторый локальный максимум этого функционала с противоположной по знаку энергией и с ортогональной конфигурацией. Заметим, что если в разложении (3) исходной матрицы минимумам соответствовали положительные веса, а максимумам – отрицательные, то возведённые в квадрат эти веса все окажутся положительными, и значит, все они будут соответствовать минимумам нового функционала, или, в силу искажающего члена R, будут иметь отрицательную энергию. Эксперимент показал, что, действительно, конфигурации как максимумов, так и минимумов исходного функционала дают на E_2 отрицательную энергию. Более того, большинство обнаруженных глубоких минимумов функционала E_2 дважды вырождены!

Поскольку наша задача состоит в том, чтобы найти глубокие минимумы исходного функционала E_1, а минимумы функционала E_2 нас, вообще говоря, не интересуют, то мы прибегли к двойному спуску. На первом этапе двойного спуска, стартуя с некоторой случайной начальной конфигурации, мы спускаемся по поверхности функционала E_2.

Дойдя до минимума, мы продолжаем спуск уже по поверхности E_1, получая в итоге минимум нашего исходного функционала.

Как показывает рис. 2, двойной спуск даёт большую плотность вероятности попадания в области глубоких минимумов. Как следствие этого, среднее значение энергии, получаемое при двойном спуске, равно – 0,894, что заметно ниже, чем среднее значение энергии при обычном спуске: – 0,841. Нормировка энергий такова, что глобальному минимуму соответствует значение энергии – 1,000.

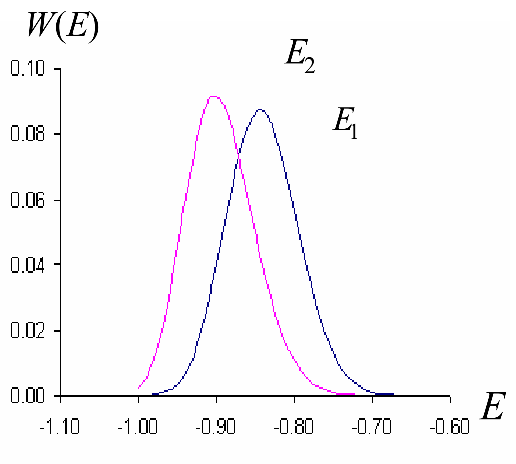


Рисунок 2 – Плотность вероятности отыскания минимума: E_1 – при простом, E_2 – двойном спусках. Энергия нормирована на глубину глобального минимума

Также двойной спуск резко уменьшил число обнаруживаемых минимумов (на миллион стартов их порядка 150 тысяч).

График распределения вероятности по расстояниям от глобального минимума представлен на рис. 3. Двойной спуск по сравнению с обычным дал увеличение вероятности попадания в глобальный минимум примерно в 160 раз.

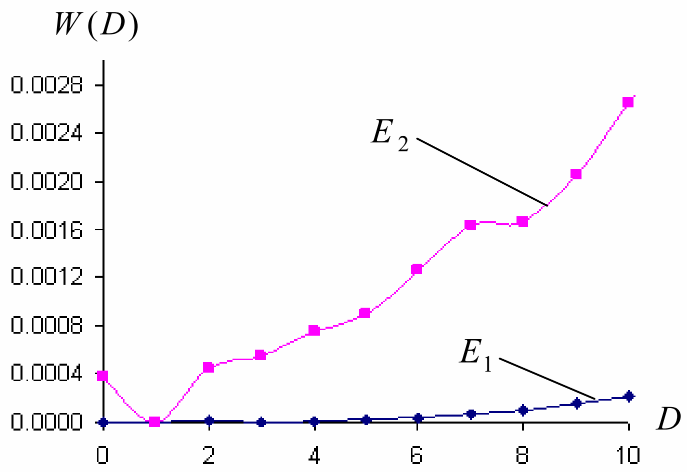


Рисунок 3 – Распределение вероятности W(D) попадания на расстояние D от глоб. минимума при: E_1 – простом спуске, E_2 – двойном. Представлены лишь расстояния от 0 до 10

Часто поиск останавливался в локальных минимумах, конфигурации которых расположены на расстоянии 2 – 3 спинов от глобального. В связи с этим мы слегка изменили нейросетевой спуск. А именно, как только нейросетевая динамика останавливалась на некотором минимуме, мы случайным образом выбирали 3 спина, переворачивали их и продолжали спуск. Для каждой стартовой конфигурации мы делали по 100 таких «отскоков», фиксируя при этом конфигурацию самого глубокого из достигнутых минимумов. На рис. 4, 5 приведены результаты для динамики с «отскоками» (для обычного и двойного спусков).

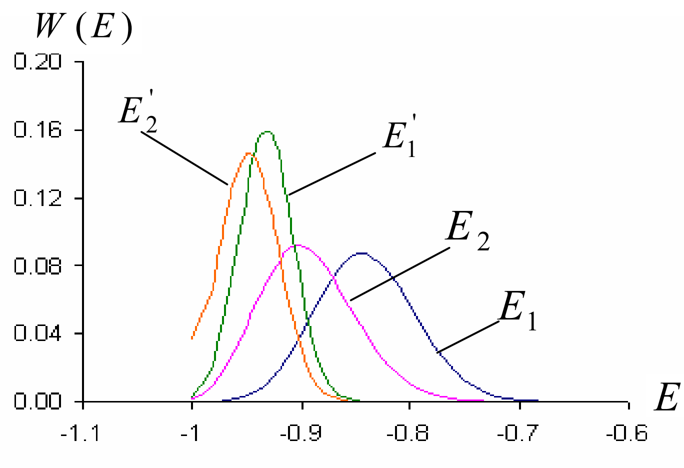


Рисунок 4 – Распределение плотности вероятности W(E) по энергии: E_1 – для простого, E_2 – двойного спусков, E'_1 – для простого спуска с отскоками, E'_2 – для двойного спуска с отскоками

При этом среднее значение энергии при простом спуске с отскоками равно – 0,932, а при двойном с отскоками – 0,948.

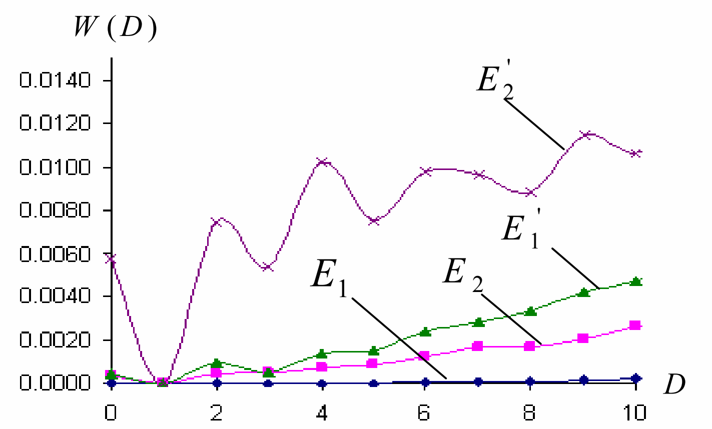


Рисунок 5 – Распределение вероятности W(D) попадания по расстояниям D от глобального минимума: E_1 – для простого, E_2 – двойного спусков, E'_1 – для простого спуска с отскоками, E'_2 – для двойного спуска с отскоками

Как оказалось, метод двойного спуска с отскоками оказался самым эффективным. На рис. 6 и 7 приведены сравнительные характеристики этого метода с простым спуском на исходном функционале.

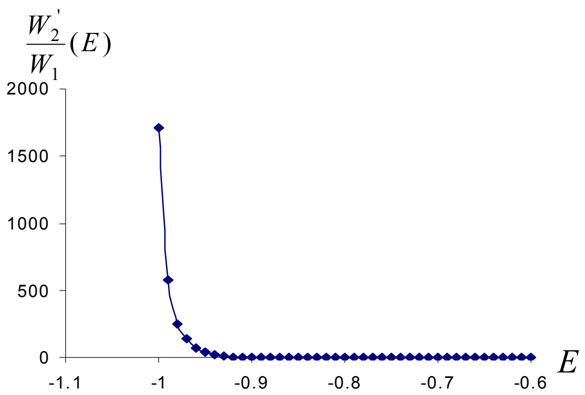


Рисунок 6 – Отношение плотностей вероятности попадания в заданный интервал энергий при двойном спуске с отскоками ($W'_2(E)$) и при простом спуске ($W_1(E)$)

Отношение вероятности попадания в глобальный минимум при двойном спуске с отскоками к вероятности попадания в него при простом спуске без отскоков ~ 2400 (рис. 7), т.е. нам удалось увеличить данный показатель на три порядка!

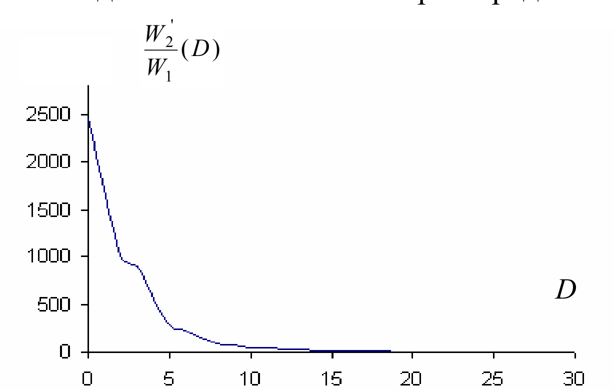


Рисунок 7 – Отношение вероятностей попадания на расстояние D от глобального минимума при двойном спуске с отскоками ($W'_2(D)$) и при простом спуске ($W_1(D)$)

Разумеется, можно было бы возводить матрицу не в квадрат, а в большую степень. При этом возведение матрицы в нечётную степень приводит к тому, что веса в 1-м члене разложения (4) тоже оказываются возведёнными в нечётную степень, а значит, сохраняют знак. Поэтому минимумы полученной матрицы будут минимумами исходной матрицы, и мы не будем тратить время на спуск с максимумов. То есть возведение матрицы в нечётную степень более эффективно, чем в чётную. Представленные в табл. 1 полученные результаты это хорошо подтверждают.

Таблица 1 – Усреднённые по 50 матрицам результаты для двойных спусков с различных степеней матриц

Обычные спуски					
Вероятность попадания:	Простой спуск (T)	Двойной спуск с матрицей T^k			
		T^2	T^3	T^4	T^5
в глоб. минимум	0,00024%	0,038%	0,2%	0,12%	0,4%
в интервал [– 0.99, – 1.00]	0,0022%	0,23%	1,1%	0,56%	2,5%
Среднее значение энергии	– 0,841	– 0,894	– 0,936	– 0,899	– 0,951
Спуски с «отскоками»					
Вероятность попадания:	Простой спуск (T)	Двойной спуск с матрицей T^k			
		T^2	T^3	T^4	T^5
в глоб. минимум	0,04%	0,67%	3,7%	1,44%	3,2%
в интервал [– 0.99, –1.00]	0,3%	3,7%	14%	5,8%	14,1%
Среднее значение энергии	– 0,932	– 0,948	– 0,975	– 0,950	– 0,974

Обсуждение результатов

Не имеет смысла возводить матрицу в степень до бесконечности, ожидая роста вероятности попадания в глубокие минимумы. Как видно из рис. 1, несмотря на то, что конфигурация глобального минимума функционала E_1 имеет меньшую энергию на функционале E_2, функционал E_2 имеет какие-то новые, ещё более глубокие, минимумы. При дальнейшем увеличении степени матрицы эти минимумы-химеры станут ещё глубже, а значит, и шире, пока полностью не затмят наш глобальный минимум. Видимо, это связано с перекрёстным членом R в разложении (4), который растёт с увеличением степени.

Из табл. 1 видно, что результаты для пятой степени матрицы мало отличаются от результатов для 3-й степени.

Следует также отметить, что при возведении матрицы в степень возникают ситуации, когда глобальный минимум вообще не находится. А именно, при обычном (без отскоков) двойном спуске с куба матрицы из 50 рассматриваемых матриц для 15 мы вообще не находили глобальный минимум. Для пятой степени число матриц с ненайденным глобальным минимумом равно 34 из 50! Ситуация для двойных спусков с отскоками несколько лучше. Однако даже в этой ситуации вероятность попадания в

узкий интервал энергий, очень близких к глобальному минимуму (от – 0,99 до – 1,00), в результате замены алгоритма случайного поиска возрастала с 0,0028% до 14,1%, т.е. более чем в 5000 раз (табл. 1). Одновременно среднее значение энергии случайно найденного минимума понизилось с величины – 0,841 до – 0,974, т.е. среднее расстояние до глобального минимума сократилось в 6 раз.

Более того, следует отметить, что даже в тех случаях, когда глобальный минимум не достигался, двойной спуск с большой вероятностью приводил систему в конфигурацию, очень близкую к глобальному минимуму: на расстоянии 2 – 5 бит от глобального минимума. Поэтому «отрицательный» результат является следствием того, что мы использовали «отскок» на три бита – наиболее простой приём выхода из мелких локальных минимумов. Однако мы не ставили перед собой цели оптимизировать выход из локального минимума – это будет сделано в дальнейшем. Нашей целью было продемонстрировать только возможность деформации потенциальной поверхности, которая приводит к существенному увеличению вероятности глобального минимума. Эта цель нами достигнута.

Литература

1. Hartmann A.K. New Optimization Algorithms in Physics / A.K. Hartmann, H. Rieger // WILEY-WCH Verlag GmbH & Co. KGaA. – Weinheim, 2004. – P. 47-71.
2. Крыжановский Б.В. Взаимосвязь глубины локального минимума и вероятности его обнаружения в обобщённой модели Хопфилда / Б.В. Крыжановский, Б.М. Магомедов, А.Л. Микаэлян // ДАН. – 2005. – Т. 405, № 3. – С. 1-5.
3. Kryzhanovsky B.V. Expansion of a matrix in terms of external products of configuration vectors / B.V. Kryzhanovsky // Optical Memory & Neural Networks (Information Optics). – 2007. – V. 16(4). – P. 187-199.

Я.М. Карандашев, Б.В. Крижановський
Ефективне збільшення області притягнення глобального мінімуму квадратичного бінарного функціонала при нейромережному пошуку
Розв'язується задача мінімізації квадратичного функціонала у конфігураційному просторі. Для ефективного збільшення області притягнення глибоких мінімумів пропонується матрицю, на якій побудований функціонал, підносити до степеня, а на отриманому новому функціоналі розв'язувати задачу мінімізації. У роботі показано на прикладі матриць двомірної спінстекольної моделі Ізінга, що така техніка приводить до зрушення спектра мінімумів у більш глибоку область, різко зменшує число знайдених мілких мінімумів і дозволяє з більшою, на 3 – 4 порядки, вірогідністю знаходити глобальний мінімум.

J.M. Karandashev, B.V. Kryzhanovsky
A Drastic Increase in the Basin of Attraction of the Global Minimum of Quadratic Binary Functional on Application of the Neural Network Search
A quadratic binary functional minimization problem is considered. To effectively increase the deep minima domains of attraction it is suggested to raise a matrix which constructed the functional on to some power, and to solve the minimization problem on the new obtained functional. By the example of matrixes of the two-dimensional Ising's model it is shown in the paper that suggested technique leads to a shift of local minima spectrum towards the region of deeper minima, reduces sharply the number of minima found, and gives an opportunity to find the global minimum with a probability on 3 – 4 orders greater.

Статья поступила в редакцию 09.06.2009.